TP248.P77.A763

STIC-ILL

From: Ly, Cheyne
Sent: Saturday, April 12, 2003 12:59 PM
To: STIC-ILL
Subject: ILL_Order

1. J Aqvist, C Medina, and JE Samuelsson

A new method for predicting binding affinity in computer-aided drug design
Protein Eng. 1994 7: 385-391.

1. MA Saqi and JM Goodfellow

Free energy changes associated with amino acid substitution in proteins
Protein Eng. 1990 3: 419-423.

C. Dune Ly
Patent Examiner
Art Unit 1631
Room 12D-04
Mailbox 12D-01
Phone: 703 308 3880
Fax: 703 749 9331

Protein Engineering vol.7 no.3 pp.385–391, 1994

A new method for predicting binding affinity in computer-aided drug design

Johan Åqvist, Carmen Medina[1] and Jan-Erik Samuelsson[1]

Department of Molecular Biology, Uppsala University, Biomedical Centre, Box 590, S-75124 Uppsala and [1]Symbicom AB, Glunten, S-75183 Uppsala, Sweden

A new semi-empirical method for calculating free energies of binding from molecular dynamics (MD) simulations is presented. It is based on standard thermodynamic cycles and on a linear approximation of polar and non-polar free energy contributions from the corresponding MD averages. The method is tested on a set of endothiapepsin inhibitors and found to give accurate results both for absolute as well as relative free energies.

Key words: binding free energies/drug design/endothiapepsin/inhibitor binding/molecular dynamics simulation

Introduction

Computational chemistry and molecular modelling have become an essential part of the modern drug design process. This type of methodology is now being used as a complement to experimental biochemical studies and 3-D structure determination by crystallographic or NMR methods. The most important applications of computational modelling in drug design comprise (i) methods for finding new 'lead compounds', (ii) interactive computer graphics for modifying and manipulating the chemical and geometrical structure of inhibitors and (iii) subsequent energy and structure refinement using molecular mechanics (MM) or dynamics (MD) calculations [for reviews, see Cohen *et al.* (1990) and Dixon (1992)]. At present, however, molecular modelling methods can only provide rather qualitative information on the affinity of different compounds for a given target site. At best, one can propose some tentative binding candidates and perhaps exclude some that do not seem to match the target site well. This may be done by considering molecular shape and electrostatic properties, but the quantitative discrimination between different ligands in terms of actual binding constants must usually be left to the experimental work of organic chemists. The reason for this is that it is extremely difficult to theoretically predict relative or absolute binding constants, i.e. free energies, for all but the very simplest cases. The only rigorous approach to this type of problem today is the so-called free energy perturbation (FEP) technique [for reviews, see Beveridge and DiCapua (1989), Jorgensen (1989) and Straatsma and McCammon (1992)] which is more or less limited to 'small perturbations', as will be discussed below.

If we consider a typical case where one wants to determine the relative free energy of binding between two compounds, A and B, the problem is described by the thermodynamic cycle

$$\begin{array}{ccccc} & & \Delta\Delta G^{w}_{sol} & & \\ & A(w) & - & B(w) & \\ \Delta G_{bind}(A) & | & & | & \Delta G_{bind}(B) \\ & A(p) & - & B(p) & \\ & & \Delta\Delta G^{p}_{sol} & & \end{array}$$

where $\Delta\Delta G^{w}_{sol}$ and $\Delta\Delta G^{p}_{sol}$ denote the differences in solvation energy between A and B in water and in the (solvated) protein site, respectively, and the ΔG_{bind}s are the corresponding binding energies; $\Delta G_{bind}(B) - \Delta G_{bind}(A) = \Delta\Delta G^{p}_{sol} - \Delta\Delta G^{w}_{sol}$. The same cycle can be used to determine the absolute binding constant of B if we let A denote a nil particle. In that case $\Delta G_{bind}(A) \approx 0$ and the binding energy is obtained as the difference between the absolute solvation energies for B in water and in the protein. With the FEP approach one calculates the free energies associated with the two unphysical paths $A(w) \rightarrow B(w)$ and $A(p) \rightarrow B(p)$, corresponding to a mutation of A into B (or the creation of B in the case where A is a nil particle). MD (or Monte Carlo) simulations are used to collect ensemble averages along the paths, which must be rather fine grained in order for the free energy to converge properly. If we consider, for example, two enzyme inhibitors, the path connecting them will involve changes in the molecular charge distribution as well as the creation/anihilation of atoms. In particular, the latter type of process converges slowly, and, thus, if the A and B 'states' are very different it will be difficult to obtain convergent free energies even with present day computer resources (Mitchell and McCammon, 1991; Pearlman and Kollman, 1991; Straatsma and McCammon, 1991). It should also be noted that most of the computing time in FEP is actually spent on uninteresting configurations that correspond to a mixture of A and B. Moreover, if large conformational changes are involved this always poses a major problem. Hence, the method only really works well when A and B are rather similar and, in fact, all applications related to drug design that have been reported fall within this category (e.g. Wong and McCammon, 1986; Bash *et al.*, 1987; Brooks, 1989; Merz and Kollman, 1989; McDonald and Brooks, 1991; Rao and Singh, 1991; Tropsha and Hermans, 1992; the largest 'perturbation' to which the FEP method has been applied seems to be a hydrogen to hexyl group transformation reported by Merz *et al.*, 1991). Drug design in practice, however, often deals with relatively large inhibitors that differ considerably from each other so that neither relative nor absolute free energies can be obtained with the FEP/MD method within reasonable computing time. This dilemma prompted us to try to devise a simplified approach that does not involve the 'mutational paths' which is what mainly hampers the FEP method.

Outline of the method

As the starting point we take the linear response approximation for electrostatic forces, which for polar solutions will yield quadratic free energy functions (potentials of mean force) in response to changes in electric fields. This is, for example, the familiar result from the Marcus' theory of electron transfer reactions (Marcus, 1964). For a system with two states, A and B, given by two potential energy functions V_A and V_B one obtains, within the approximation of harmonic free energy functions of equal curvature, the relationship [see Appendix and for example, Lee *et al.* (1992) and references therein]

$$\lambda = \langle V_B - V_A \rangle_A - \Delta G_{AB} = \langle V_A - V_B \rangle_B + \Delta G_{AB} \quad (1)$$

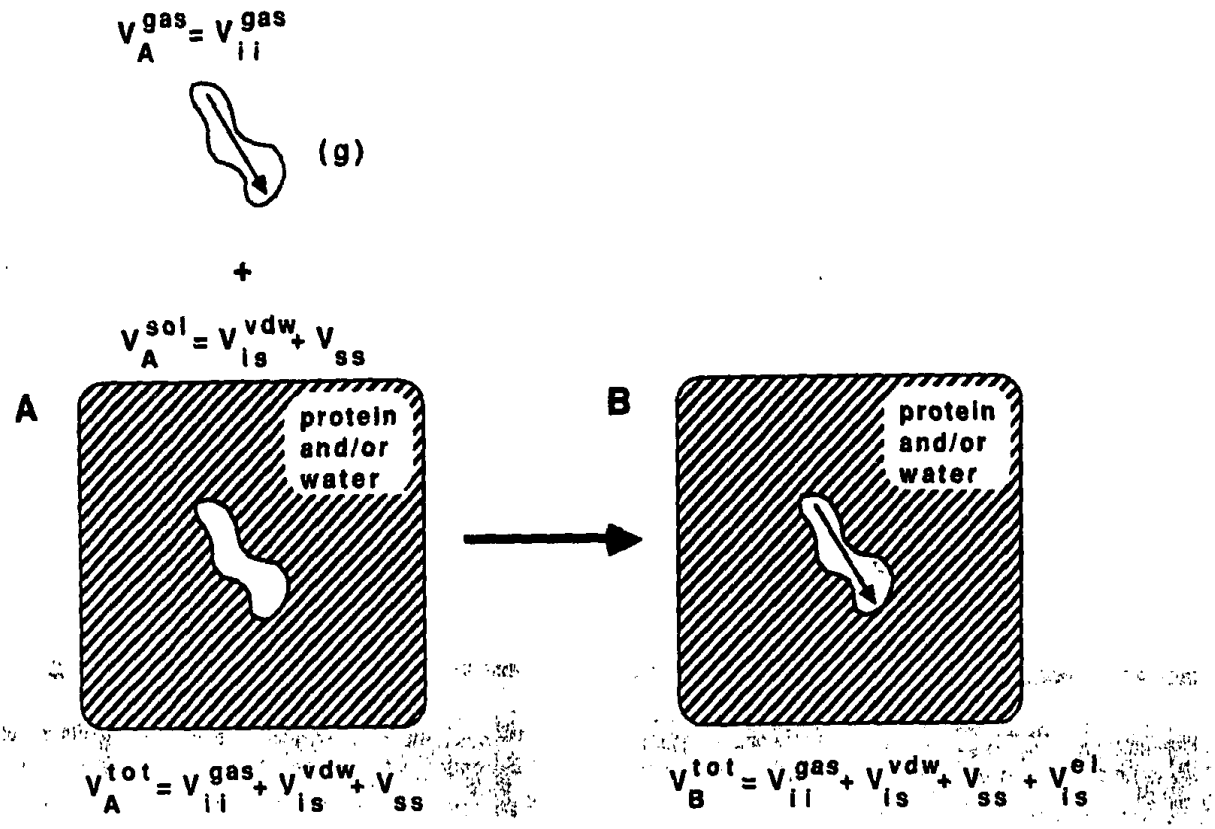

Fig. 1. Illustration of how the approximation of equation (2) can be used to estimate the electrostatic contribution to ΔG_{sol} in a given environment, that can either be water or protein. The state *A* corresponds to a system with the isolated inhibitor in the gas phase and a ready made van der Waals cavity in the condensed system. State *B* is simply the solvated inhibitor in water or in the protein's active site. These states are given by the two potentials V_A and V_B. For any given configuration of the system, equation (2) will reduce to $\frac{1}{2} V_{i-s}^{el}$ since the solvent configuration in state *A* is uncorrelated with the charge distribution of the solute in state *B*.

where ΔG_{AB} is the free energy difference between *B* and *A*, λ the corresponding reorganization energy and $\langle\rangle_i$ denotes a mean evaluated near the minimum of the potential *i*. We thus have

$$\Delta G_{AB} = \tfrac{1}{2} (\langle \Delta V \rangle_A + \langle \Delta V \rangle_B) \quad (2)$$

where ΔV now denotes the energy difference $V_B - V_A$. If we consider the hydration of a single ion this can be shown to give $\Delta G_{sol}^{el} = \frac{1}{2} \langle V_{i-s}^{el} \rangle$, i.e. that the electrostatic contribution to the solvation energy equals half of the corresponding ion–solvent interaction energy (Warshel and Russell, 1984; Roux *et al.*, 1990). Returning now to our inhibitor binding problem we can exploit this result as indicated in Figure 1. For each solvation process, i.e. solvation of the inhibitor in water and inside the protein, we consider two states where the first has the inhibitor molecule *in vacuo* and a non-polar cavity (given, for example, by a Lennard–Jones potential) already made in the given environment. The second state corresponds to the intact inhibitor molecule surrounded by water or the solvated protein. The linear response approximation will then again give us $\Delta G_{sol}^{el} = \frac{1}{2} \langle V_{i-s}^{el} \rangle$ where V_{i-s}^{el} is the solute–solvent electrostatic term. Hence, the electrostatic contribution to the binding free energy can be approximated by $\Delta G_{bind}^{el} \simeq \frac{1}{2} \langle \Delta V_{i-s}^{el} \rangle$ (where the Δ now refers to the difference between protein and water) which is obtained from two MD simulations of the solvated inhibitor and of the inhibitor–protein complex.

The validity of the linear response results in the case of ionic solvation has been confirmed, for example, in the study by Roux *et al.* (1990). We have also performed some additional calculations on simple systems that corroborate the approximation of equation (2). Tests were, for example, carried out by comparing the free energy obtained from FEP/MD simulations of charging Na^+ and Ca^{2+} ions in a spherical water system (Åqvist, 1990) with the corresponding $\langle V_{i-s}^{el} \rangle$ from 75 ps MD trajectories. This yielded factors relating ΔG_{sol}^{el} to $\langle V_{i-s}^{el} \rangle$ of 0.49 for Na^+ and 0.52 for Ca^{2+}, both values being close to the predicted result of ½. A similar test on the charging of a methanol molecule, given by the OPLS potential (Jorgensen, 1986), in water gave a $\Delta G_{sol}^{el} : \langle V_{i-s}^{el} \rangle$ ratio of 0.43. Some calculations were also done on rigid two-atom dipoles with the charge centres separated by 4 Å. For a $Q = \pm 2$ (elementary charge units) dipole a ratio of 0.53 was obtained while $Q = \pm 1$ and $Q = \pm 0.25$ gave ratios of 0.47 and 0.49 respectively. Charging a sodium ion given by the parameters in Åqvist (1990) to +0.25 charge units gave a ratio of $\Delta G_{sol}^{el} : \langle V_{i-s}^{el} \rangle = 0.55$. Hence, the approximation of equation (2) seems to be reasonable although it might be somewhat less accurate for very weak fields. Whether this is due to non-linear effects or simply to the fact that the relative fluctuations of V_{i-s}^{el} are larger is not entirely clear and deserves further study.

Compared to the electrostatic problem above it is a more difficult question how to account for the contribution of non-polar interactions and hydrophobic effects to the free energy of binding, which we will call ΔG_{bind}^{vdw}. What is desirable is to be able to estimate this contribution from the non-polar (or van der Waals) interaction energies. The liquid theories of Chandler and coworkers (Pratt and Chandler, 1977; Chandler *et al.*, 1983) have been successfully used to analyse hydrophobic effects and to

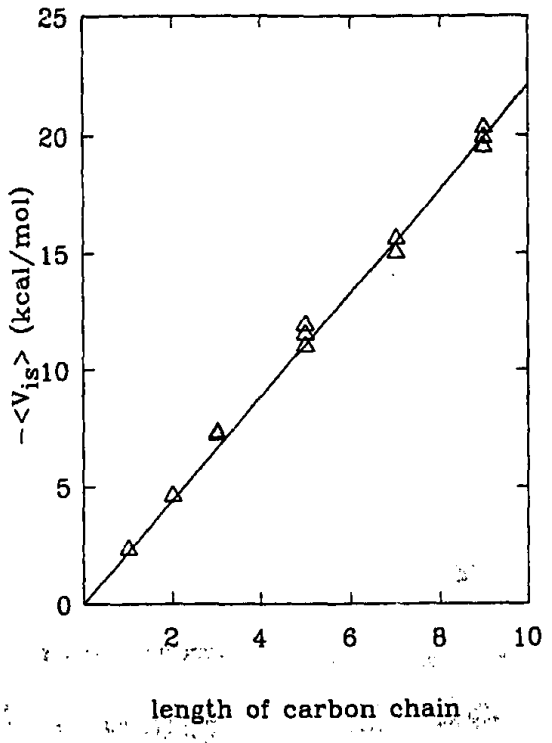

Fig. 2. Calculated dependence of the mean solute–solvent interaction energy on the length of the carbon chain for *n*-alkanes solvated in water.

calculate free energies of transfer for some non-polar molecules (Pratt and Chandler, 1977), but no analytical treatment of that kind seems possible for solvation in an inhomogenous environment such as a protein's active site. It can, however, be noted that the experimental free energy of solvation for various hydrocarbon compounds, such as *n*-alkanes, depends approximately linearly on the length of the carbon chain both in their own liquids as well as in water (Ben-Naim and Marcus, 1984). We have carried out MD simulations of *n*-alkanes solvated in water (Figure 2) and in a non-polar van der Waals solvent (not shown) which indicate that the mean solute–solvent interaction energies also vary approximately linearly with the number of carbons in the chain (the relationships being different in different solvents, of course). It would thus seem possible that a simple linear approximation of ΔG^{vdw}_{bind} from $\langle \Delta V^{vdw}_{i-s} \rangle$ might be able to account for the non-polar binding contribution. For instance, if we let σ be some appropriate measure of the size of the solute and if the solute–solvent van der Waals interaction energies and the corresponding non-polar free energy contributions (both in water and protein) depend linearly on σ, such that, $\langle V^{vdw}_p \rangle = \alpha_p \sigma$, $\langle V^{vdw}_w \rangle = \alpha_w \sigma$, $\Delta G^{vdw}_p = \beta_p \sigma$ and $\Delta G^{vdw}_w = \beta_w \sigma$, then we obtain $\Delta G^{vdw}_{bind} = \frac{\beta_p - \beta_w}{\alpha_p - \alpha_w} \langle \Delta V^{vdw}_{i-s} \rangle$. Since it seems difficult to derive a factor relating the two quantities in a reliable way from purely theoretical considerations, we take the approach here to empirically try to determine such a relationship that is capable of reproducing experimental binding data. Thus, we will approximate the free energy of binding by

$$\Delta G_{bind} = \tfrac{1}{2} \langle \Delta V^{el}_{i-s} \rangle + \alpha \langle \Delta V^{vdw}_{i-s} \rangle \qquad (3)$$

and attempt to determine the parameter α by empirical calibration.

As a test system we have chosen endothiapepsin (EP) belonging to the family of aspartic proteinases (see, e.g. Fruton, 1976; Davies, 1990), a class of enzymes for which numerous studies of inhibitor binding have been reported. Crystal structures of native endothiapepsin and inhibitor complexes have been published by Blundell and coworkers (Foundling *et al.*, 1987; Veerapandian *et al.*, 1990). In this work we use the crystal structure of endothiapepsin in complex with the inhibitor H218/54 (Figure 3, I_1), recently determined by Symbicom AB (D.Ogg *et al.*, in preparation), as the structural starting point for our computations. Four other inhibitor compounds were used in this work and their chemical structures are also shown in Figure 3. For simplicity we label these molecules $I_1, \ldots, I_5$. Experimental binding data for the inhibitors have been obtained by Falknäs and Deinum (A.Hässle, unpublished results).

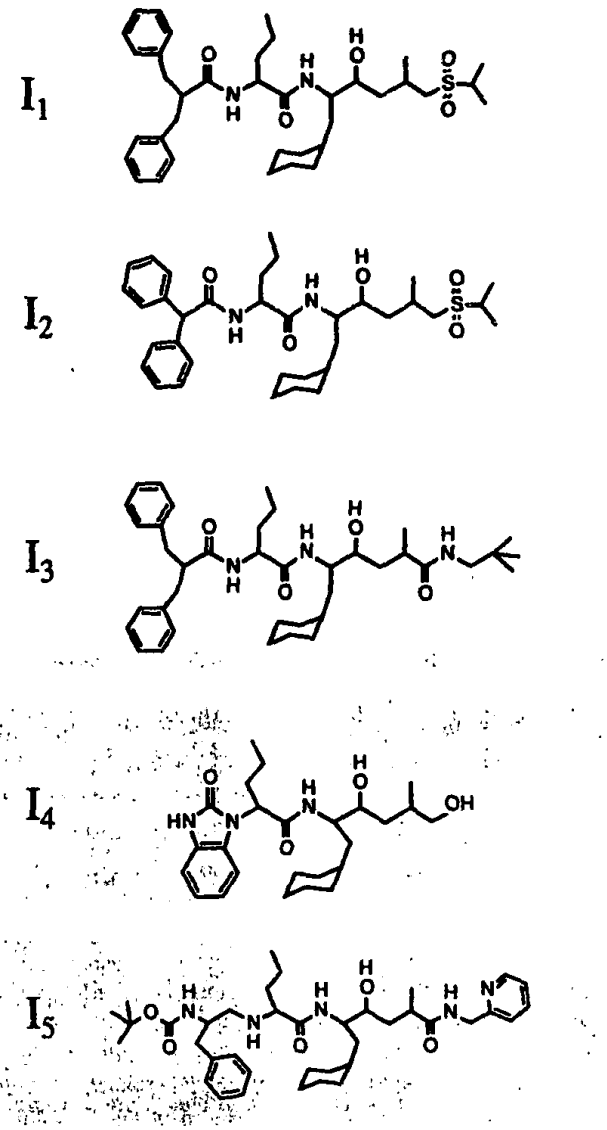

Fig. 3. Chemical structures of the inhibitors $I_1, \ldots, I_5$ used in the calculations.

Computational details

Starting from the experimental structure of the EP–I_1 complex the other inhibitors were manually built into the EP active site using the FRODO (Jones, 1978) and Hydra (Hubbard, 1984) graphics software. This model building appears relatively straightforward since all the compounds contain the invariant

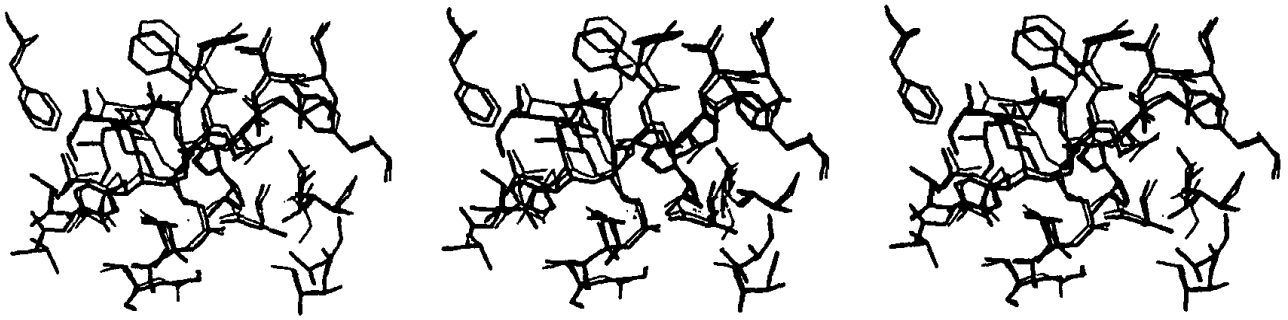

Fig. 4. *Stereo view of the 50 ps mean MD structure of the EP-I_1 complex (thin lines) superimposed on the corresponding crystallographic structure (D.Ogg et al., in preparation). The active site of the enzyme is shown with the bound inhibitor in the centre of the picture.*

hydroxyethylene transition state isostere adjacent to a peptide group, which was also present in the inhibitors studied by Blundell's group. After initial energy minimization, MD simulations consisting of 25 ps of equilibration and 50 ps of data collection were performed for each of the inhibitors, solvated in water as well as bound to the solvated protein.

A modified version (J.Aqvist, unpublished) of the ENZYMIX program (Warshel and Creighton, 1989) was used for all MD simulations together with the GROMOS potential (van Gunsteren and Berendsen, 1987). This force field was revised by us with respect to the oxygen-extended carbon (CH_n) interactions, since the standard GROMOS parameters were found from FEP simulations to give an incorrect value of $\Delta G_{hydr} = +0.1 \pm 0.5$ kcal/mol for methane (J.Åqvist, unpublished results). By changing the oxygen repulsive Lennard–Jones 6/12 parameter (A_O) for $O-CH_n$ interactions from 421.0 to 793.3 (kcal/mol) ½ Å^6, ΔG_{hydr} for methane becomes 2.5 ± 0.4 kcal/mol, in better agreement with the experimental value of 2.0 kcal/mol (Ben-Naim and Marcus, 1984). Note that this revision is quite important for a correct description of the hydrophobic effect. Also the Lennard–Jones parameters for charged carboxylate groups have been revised earlier in order to better reproduce experimental solvation energies (Åqvist *et al.*, 1993). The present calculations used a spherical water droplet of radius 16 Å for the simulations of inhibitors in solution and a corresponding 16 Å sphere containing both protein atoms and water in the simulations of bound inhibitors. The atoms within this sphere were free to move while protein atoms beyond 16 Å were restrained to their crystallographic positions. An interaction cut-off radius of 8 Å was used, except for interactions involving the inhibitor and the MD time step was 0.001 ps. The equilibration phase of the protein simulations consisted of 5 ps of successive heating of the system and weakening of harmonic positional restraints that were applied to the protein atoms. After this period all restraints within the 16 Å sphere were set to zero and the system was further equilibrated at the final temperature of 298 K for another 20 ps. The r.m.s. coordinate deviation for the inhibitor atoms between the following 50 ps average MD structure and the experimental EP–I_1 complex is 0.94 Å. Figure 4 shows a superposition of these two structures and it can be seen that the agreement is quite satisfactory.

Results and discussion

We used the first four inhibitors ($I_1, \ldots, I_4$) as our calibration set in order to determine the optimal value of the parameter α and examine the success of equation (3) in reproducing the experimental results. Table I shows the observed and calculated absolute free energies of binding for the inhibitors $I_1, \ldots, I_4$ using the value of $\alpha = 0.161$ which optimizes the r.m.s.

Table I. Calculated and observed (Falknäs and Deinum, unpublished results) absolute binding free energies for inhibitors $I_1, \ldots, I_4$ to endothiapepsin (in kcal/mol)

I_x	ΔG^{calc}_{bind}	ΔG^{obsd}_{bind}
1	−11.16	−10.69
2	−8.23	−7.93
3	−11.14	−11.67
4	−6.29	−6.57

Table II. Calculated and observed relative binding free energies for inhibitors $I_1, \ldots, I_4$ to endothiapepsin (in kcal/mol).

I_x-I_y	$\Delta\Delta G^{calc}_{bind}$	$\Delta\Delta G^{obsd}_{bind}$
1–2	−2.92	−2.76
1–3	−0.01	+0.98
1–4	−4.87	−4.12
2–3	+2.91	+3.74
2–4	−1.95	−1.36
3–4	−4.86	−5.10

agreement with the experiment. This value of α gives a mean unsigned error of 0.39 kcal/mol for the calibration set and the largest error is 0.53 kcal/mol for inhibitor I_3. Initially we had not expected such an accurate fit for the absolute binding energies but had mainly hoped to be able to apply and calibrate equation (3) with respect to the relative ones. If the calibration instead is done using relative binding energies one obtains a very similar value of α (0.169) with mean unsigned errors that are virtually identical to those above. This result is quite remarkable and, thus, lends further support to the approximation of equation (3). The calculated and observed relative free energies of binding (using $\alpha = 0.161$) are shown in Table II and the mean unsigned error is 0.59 kcal/mol. The largest error here is 0.99 kcal/mol for the I_1/I_3 selectivity, but all other relative free energies have the correct sign and are within 0.8 kcal/mol of the experimental result. It is particularly interesting to note that the calculations are able to discriminate the low-affinity inhibitor I_4 quite well from the high-affinity ones (I_1 and I_3).

Encouraged by the above results, we wanted to try to assess the predictive power of our approach by modelling an inhibitor not present in the calibration set. For this we chose the inhibitor I_5 which, as can be seen from Figure 2, differs significantly in its chemical structure from any member of the calibration set. This molecule was built into the EP active site and subjected to the same simulation procedure as the other inhibitors. The predicted absolute free energy of binding for I_5 is −9.70 kcal/mol which is in excellent agreement with the

Table III. Calculated and observed relative binding free energies involving inhibitor I_5 (in kcal/mol)

$I_x - I_y$	$\Delta\Delta G_{bind}^{calc}$	$\Delta\Delta G_{bind}^{obsd}$
1-5	-1.46	-0.85
2-5	+1.47	+1.91
3-5	-1.44	-1.83
4-5	+3.41	+3.27

corresponding observed result of $\Delta G_{bind}(I_5) = -9.84$ kcal/mol. The calculated relative binding free energies with respect to the four inhibitors in the calibration set are given in Table III, where it can be seen that all the pairwise selectivities involving I_5 are correctly predicted by our simulations, the maximum error in this case being 0.61 kcal/mol. The above results indicate that the simple linear approximation of the free energy given by equation (3) is able to describe the main physics (electrostatic and hydrophobic interactions) of the binding process in a satisfactory way. (In retrospect, it is also of interest to see what the effect would be of choosing different inhibitor subsets for the calibration of α. It then turns out that the possible subsets of four inhibitors give values in the range $0.158 \leq \alpha \leq 0.165$. For example, $\alpha = 0.165$ is obtained if I_1 is left out of the calibration and its predicted binding energy then becomes -11.31 kcal/mol. If all five inhibitors are included in the calibration one obtains a value of $\alpha = 0.162$.)

It is, of course, important to try to identify and separate the different types of errors involved in the present method and suggest how they possibly could be dealt with. There are basically four sources of errors, namely (i) inaccurate starting structures due to incorrect model building, (ii) possible deficiencies in potential energy functions, (iii) poor MD convergence, for example, due to short trajectories and (iv) the approximation of equation (3) itself. The only remedy for errors of the first type is to obtain as much structural information as possible and in order to assess their magnitude it would be desirable to carry out calculations on a set of inhibitor complexes whose 3-D structures have all been experimentally determined. Although MM/MD potentials are continuously being refined by many research groups errors of the second type may still be considerable. It is, for example, not yet entirely clear how well customary protein force fields actually reproduce relevant energetic properties. In our case we can note that the force field used here has required some revision in order to reproduce essential solvation properties (see above).

The convergence properties of MD simulations depend on how far from equilibrium the initial structure is, but judging from the present study it seems that one can reach satisfactory convergence within reasonable computing time. For example, by comparing means over the first and second halves of the MD trajectories we obtain mean (over all five inhibitors) errors of ± 0.35 and ± 0.75 kcal/mol for V_{i-s}^{vdw} and V_{i-s}^{el} respectively, in the protein; the corresponding errors in water are ± 0.46 and ± 0.62 kcal/mol. This would yield a nominal error range of ± 0.82 kcal/mol in equation (3) originating from the MD convergence uncertainty. Here, it can be noted that an advantage with the present approach is probably that it focuses on simulation of the thermodynamically relevant states compared, for example, to FEP calculations where most of the computing time is spent on the paths between such states. The errors associated with the approximation of equation (3) are at the present stage difficult to estimate quantitatively although the agreement with experimental binding data obtained here indicates that the linear approximation is reasonable. A larger calibration set would obviously be desirable but this may be regarded as a matter of refinement, which we leave for future work. As mentioned above, Ben-Naim and Marcus (1984) have shown for several classes of hydrocarbon chain containing organic compounds that a linear fit of $\Delta G_{sol} = kn + l$, where n is the number of carbon atoms in the chain, is quite accurate both in water and non-polar solvents. It can, however, be noted that in some cases the extrapolation, l, of ΔG_{sol} to zero chain length is non-zero (e.g. for n-alkanes in water $l = 1.42$ kcal/mol). This might suggest that an additional constant term should be added to equation (3), reflecting a difference in extrapolation to 'zero inhibitor size' (between the water and protein environments) of the non-polar contribution to ΔG_{sol}. However, such a term would again be difficult to derive in a reliable way for the systems we are dealing with and it does not seem warranted at the present stage to introduce another empirical parameter in equation (3) (in fact, doing so does not improve the fit reported in Table I significantly). It is also important to emphasize here that equation (3) with the above parametrization of α is not an equation for the individual solvation energy terms, since the factor α represents the combined effect of several energy/free energy relations [however, one could of course try to parametrize equation (3) for solvation energies which is of interest for calculating free energies of transfer between different solvents]. The fact that we always deal with solvation energy differences may also cause some cancellation of possible systematic errors. Furthermore, due to the larger weight given by equation (3) to the ΔV_{i-s}^{el} term one might perhaps expect the formula to give better results for polar inhibitors, as long as the electrostatic linear response approximation holds. When comparing inhibitors of different charge states long-range corrections of the Born type (see, e.g. Straatsma and Berendsen, 1988; Åqvist, 1990) would obviously become important and may make accurate predictions more difficult, although counter ions may be used to keep the system overall neutral.

It also remains to be seen whether the empirical parameter α is readily transferable to other systems or whether it will display some system dependency. One thing that is clear, given the fact that the parametrization of force fields can differ considerably, is that α is likely to be force field-specific. If we are unfortunate this parameter might also, for example, reflect the nature of the particular protein site. But, even if this turns out to be the case we feel that the present approach could be quite efficient for discriminating between a large number of potential inhibitors to a given target site, thereby reducing the need to synthesize all of them. One would then, of course, have to go through an initial round of calibration but this might be worth the effort. It should also be emphasized that there does not really exist any reliable computational method today that could match the quantitative level of results reached here for such large and widely differing inhibitors.

Acknowledgements

We thank Dr J.Deinum (Astra Hassle AB) for providing the data on inhibitor binding to endothiapepsin and Dr Derek Ogg (Symbicom AB) for providing the structure of the enzyme–inhibitor complex. Support to J.Å. from the Swedish Natural Science Research Council (NFR) is gratefully acknowledged.

References

Åqvist,J. (1990) *J. Phys. Chem.*, **94**, 8021–8024.

Åqvist,J., Fothergill,M. and Warshel,A. (1993) *J. Am. Chem. Soc.*, **115**, 631–635.

Bash,P.A., Singh,U.C., Brown,F.K., Langridge,R. and Kollman,P.A. (1987) *Science*, **235**, 574–576.

Ben-Naim,A. and Marcus,Y. (1984) *J. Chem. Phys.*, **81**, 2016–2027.
Beveridge,D.L. and DiCapua,F.M. (1989) *Annu. Rev. Biophys. Biophys. Chem.*, **18**, 431–492.
Brooks,C.L.,III (1989) In van Gunsteren,W.F. and Weiner,P.K. (eds), *Computer Simulation of Biomolecular Systems*. ESCOM, Leiden, pp. 73–88.
Chandler,D., Weeks,J.D. and Andersen,H.C. (1983) *Science*, **220**, 787–794.
Cohen,N.C., Blaney,J.M., Humblet,C., Gund,P. and Barry,D.C. (1990) *J. Med. Chem.*, **33**, 883–894.
Davies,D.R. (1990) *Annu. Rev. Biophys. Biophys. Chem.*, **19**, 189–215.
Dixon,J.S. (1992) *TIBTECH*, **10**, 357–363.
Foundling,S.I., Cooper,J., Watson,F.E., Cleasby,A., Pearl,L.H., Sibanda,B.L., Hemmings,A., Wood,S.P., Blundell,T.L., Valler,M.J., Norey,C.G., Kay,J., Boger,J., Dunn,B.M., Leckie,B.J., Jones,D.M., Atrash,B., Hallett,A. and Szelke,M. (1987) *Nature*, **327**, 349–352.
Fruton,J.S. (1976) *Adv. Enzymol.*, **44**, 1–36.
Hubbard,R.E. (1984) In *The Representation of Protein Structure, Proc. Computer-Aided Molecular Design Conference*, Oyez, New York, p. 99.
Jones,T.A. (1978) *J. Appl. Crystal.*, **11**, 268–272.
Jorgensen,W.L. (1986) *J. Phys. Chem.*, **90**, 1276–1284.
Jorgensen,W.L. (1989) *Acc. Chem. Res.*, **22**, 184–189.
King,G. and Warshel,A. (1990) *J. Chem. Phys.*, **93**, 8682–8692.
Lee,F.S., Chu,Z.-T., Bolger,M.B. and Warshel,A. (1992) *Protein Engng*, **5**, 215–228.
Marcus,R.A. (1964) *Ann. Rev. Phys. Chem.*, **15**, 155–196.
McDonald,J.J. and Brooks,C.L.,III (1991) *J. Am. Chem. Soc.*, **113**, 2295–2301.
Merz,K.M.,Jr and Kollman,P.A. (1989) *J. Am. Chem. Soc.*, **111**, 5649–5658.
Merz,K.M.,Jr, Murcko,M.A. and Kollman,P.A. (1991) *J. Am. Chem. Soc.*, **113**, 4484–4490.
Mitchell,M.J. and McCammon,J.A. (1991) *J. Comput. Chem.*, **12**, 271–275.
Pearlman,D.A. and Kollman,P.A. (1991) *J. Chem. Phys.*, **94**, 4532–4545.
Pratt,L.R. and Chandler,D. (1977) *J. Chem. Phys.*, **67**, 3683–3704.
Rao,B.G. and Singh,U.C. (1991) *J. Am. Chem. Soc.*, **113**, 6735–6750.
Roux,B., Yu,H.-A. and Karplus,M. (1990) *J. Phys. Chem.*, **94**, 4683–4688.
Straatsma,T.P. and Berendsen,H.J.C. (1988) *J. Chem. Phys.*, **89**, 5876–5886.
Straatsma,T.P. and McCammon,J.A. (1991) *J. Chem. Phys.*, **95**, 1175–1188.
Straatsma,T.P. and McCammon,J.A. (1992) *Ann. Rev. Phys. Chem.*, **43**, 407–435.
Tropsha,A. and Hermans,J. (1992) *Protein Engng*, **5**, 29–33.
van Gunsteren,W.F. and Berendsen,H.J.C. (1987) *Groningen Molecular Simulation (GROMOS) Library Manual*. Biomos BV, Nijenborgh 16, Groningen, The Netherlands.
Veerapandian,B., Cooper,J.B., Sali,A. and Blundell,T.L. (1990) *J. Mol. Biol.*, **216**, 1017–1029.
Warshel,A. and Russell,S.T. (1984) *Q. Rev. Biophys.*, **17**, 283–422.
Warshel,A. and Creighton,S. (1989) In van Gunsteren,W.F. and Weiner,P.K. (eds), *Computer Simulation of Biomolecular Systems*. ESCOM, Leiden, pp. 120–138.
Wong,C.F. and McCammon,J.A. (1986) *J. Am. Chem. Soc.*, **108**, 3830–3832.
Zwanzig,R.W. (1954) *J. Chem. Phys.*, **22**, 1420–1426.

Received June 28, 1993; revised November 4, 1993; accepted November 16, 1993

Appendix

Figure 5 depicts schematically the free energy functions for a system with two states defined by potential surfaces V_A and V_B. The free energy functions are defined as the potential of mean force with respect to some reaction coordinate, X, that is, the probability distribution for X on the potential surface V_A is

$$p_A(X) = <\delta(X'-X)>_A = \frac{\int \delta(X'-X)e^{-\beta V_A(\Gamma)}\,d\Gamma}{\int e^{-\beta V_A(\Gamma)}\,d\Gamma} \qquad (4)$$

and the corresponding free energy function is then defined as

$$\Delta g_A(X) = -\beta^{-1} \ln p_A(X) \qquad (5)$$

where $\beta^{-1} = k_B T$. Analogously, the free energy function on the V_B surface is written as

$$\Delta g_B(X) = -\beta^{-1} \ln p_B(X) + \Delta G_{AB} \qquad (6)$$

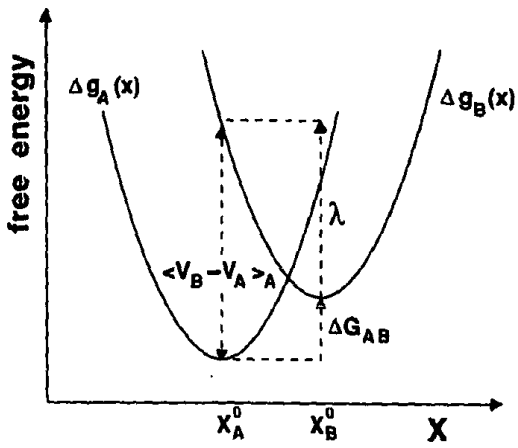

Fig. 5. Schematic free energy curves for a system obeying the linear response approximation. See text for details.

where the constant ΔG_{AB} is the (reaction) free energy associated with changing the state of the system from A to B. In the present case, we consider this change to only involve electrostatic variables so that the difference between V_A and V_B reduces to coulombic terms. The reaction coordinate, X, describes the change from state A to state B and can be chosen in different ways (see, e.g. King and Warshel, 1990), for instance, as the energy gap between the two surfaces or as a suitable charging parameter. The essence of the linear response approximation is that the free energy functions assume parabolic form and are characterized by the same force constant (i.e. the system responds linearly to electrostatic forces with a certain dielectric constant). This is the standard dielectric continuum result of which Marcus' ET theory provides a nice illustration (Marcus, 1964). Equation (1), which relates the reorganization energy (λ) to the reaction free energy and mean energy gap, then follows from the simple geometric relationship between the two free energy curves. It should thus be noted that the factor of ½ in equation (2), that can be derived in different ways (e.g. Warshel and Russell, 1984; Roux *et al.*, 1990), originates from the assumption of harmonic free energy curves of equal curvature. If the curvatures of the parabolas are not equal one obtains an additional term of $-\frac{1}{2}\Delta\lambda$ in equation (2), where the difference in reorganization energy between the two states is related to the origin shift and force constants of the parabolas by $\Delta\lambda = (\Delta X^0)^2(k_B - k_A)$.

An alternative derivation of equation (2) can be carried out starting from the perturbation expression for the free energy difference, which can then be written as (Zwanzig, 1954)

$$\Delta G = -\beta^{-1} \ln <e^{-\beta(V_B - V_A)}>_A \qquad (7)$$

Expansion of the exponent and logarithm then yields

$$\begin{aligned}
\Delta G &= -\beta^{-1} \ln <1-\beta V_B - V_A) + (\tfrac{\beta^2}{2}) \\
&\qquad (V_B-V_A)^2 - \ldots>_A \\
&= -\beta^{-1} \ln \{1-\beta <(V_B-V_A)>_A + (\tfrac{\beta^2}{2}) \\
&\qquad <(V_B-V_A)^2>_A - \ldots\} \\
&= -\beta^{-1} \{-\beta <V_B-V_A>_A + (\tfrac{\beta^2}{2}) \\
&\qquad <(V_B-V_A)^2>_A \\
&\quad - \tfrac{1}{2}(-\beta < V_B-V_A>_A + (\tfrac{\beta^2}{2}) \\
&\qquad <(V_B-V_A)^2>_A)^2 + \ldots\}
\end{aligned} \qquad (8)$$

which can be rearranged as

$$\Delta G = \langle \Delta V \rangle_A - (\tfrac{\beta}{2}) \langle (\Delta V - \langle \Delta V \rangle_A)^2 \rangle_A + (\tfrac{\beta^2}{6}) \langle (\Delta V - \langle \Delta V \rangle_A)^3 \rangle_A - (\tfrac{\beta^3}{24}) [\langle (\Delta V - \langle \Delta V \rangle_A)^4 \rangle_A - 3 \langle (\Delta V - \langle \Delta V \rangle_A)^2 \rangle_A^2] + \ldots \quad (9)$$

where $\Delta V = V_B - V_A$. In the same way we obtain by averaging on the potential surface V_B

$$\Delta G = \langle \Delta V \rangle_B + (\tfrac{\beta}{2}) \langle (\Delta V - \langle \Delta V \rangle_B)^2 \rangle_B + \ldots \quad (10)$$

Adding the two equations then gives us

$$\Delta G = \tfrac{1}{2} [\langle \Delta V \rangle_A + \langle \Delta V \rangle_B] - (\tfrac{\beta}{4}) [\langle (\Delta V - \langle \Delta V \rangle_A)^2 \rangle_A - \langle (\Delta V - \langle \Delta V \rangle_B)^2 \rangle_B] + \ldots \quad (11)$$

Although equation (2) could now be retained by just dropping the terms of second and higher order, it should be pointed out that a simple truncation to first order in ΔV would correspond to the approximations $\Delta G \simeq \langle \Delta V \rangle_A$ (from equation 9) and $\Delta G \simeq \langle \Delta V \rangle_B$ (from equation 10), that is, an assumption of the mean energy gaps being equal on the two surfaces, which is clearly not justified. Therefore, the form of equation (2) is formally obtained by truncating terms of third order and higher and by assuming that the mean square fluctuations of the energy gap on the two surfaces are equal and, thus, will cancel. Hence, the approximations here are identical to those behind Figure 5, that is, truncation of the energy gap expansions beyond the second moment is equivalent to the assumption of harmonic free energy functions and the cancellation of the fluctuation terms is equivalent to the assumption of equal force constants for the two states (the proof of this is straightforward). In order for equation (2) to be valid it is, actually, sufficient with the somewhat weaker assumption of symmetric free energy functions with equal expansion coefficients. It can also be noted that equations (9) and (10) provide a means of obtaining reorganization energies from simulations and, correspondingly, in cases where the equal curvature approximation does not hold equation (11) provides a means for evaluating the difference between the force constants and, thus, for the difference in reorganization energy.